DOCUMENT RESUME

ED 266 636 FL 015 450

AUTHOR Perkins, Kyle; And Others
TITLE Evidence for Unique Articulatory and Auditory Dimensions in ESL Data.
PUB DATE 85
NOTE 16p.; In: Practice and Problems in Language Testing 8. Papers presented at the International Language Testing Symposium of the Interuniversitare Sprachtestgruppe (IUS) (8th, Tampere, Finland, November 17-18, 1984); see FL 015 442.
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150) -- Statistical Data (110)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Adult Learning; *Data Interpretation; *English (Second Language); Foreign Countries; Higher Education; *Language Proficiency; *Language Tests; *Linguistic Theory; Listening Comprehension; Second Language Learning

ABSTRACT

A study was carried on to analyze a data set in order to determine whether evidence can be found to support or refute the unitary factor hypothesis of language proficiency, which asserts that there is a single language dimension or factor underlying different language tests. The subjects were 50 advanced adult learners of English as a second language with a variety of native languages. Data were elicited with the Wepman Auditory Discrimination Test, the Seashore Measure of Musical Talents tests of pitch and loudness, and the Test of Spoken English (TSE) for oral proficiency. Statistical analysis of the results revealed that there are two distinct factors, auditory and articulatory, present in the data and suggested that more sophisticated analysis of additional data might reveal evidence of a method factor, disconfirming the unitary competence hypothesis and supporting multiple-factor models. Appended are the TSE rating schedules and tables of statistical data. (MSE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

BEST COPY AVAILABLE

ED266636

EVIDENCE FOR UNIQUE ARTICULATORY AND AUDITORY DIMENSIONS IN ESL DATA

Kyle Perkins, Sheila R. Brutten, and Paul J. Angelis (Southern Illinois University)

1 Introduction

In language testing, unidimensionality refers to the assumption that subjects' responses to test items are a function of one and only one construct measured by the test under consideration. On the other hand, multidimensionality refers to the assumption that item responses are a function of more than one construct. The notion of dimensionality has been extended in recent years to include investigations of the construct validity of language tests and the hypothesis of a single language proficiency dimension or factor underlying different language tests.

Several investigators have employed principal components analysis in their analysis of test data and have reported their findings in support of the single unitary factor hypothesis--also known as the unitary competence hypothesis (e.g., Oller 1979; Oller and Hinofotis 1980). On the other hand, Bachman and Palmer (1981) refute the unitary factor hypothesis, instead presenting evidence for a model with two correlated trait factors and three method factors. In addition, Upshur and Homburg (1983) present evidence in favor of multiple-factor models of language proficiency.

The purpose of this investigation is to analyze a data set to determine whether evidence can be found to support or to refute the unitary factor hypothesis of language proficiency.

2 Methodology

The subjects for this investigation were 50 adult students enrolled in the most advanced full-time level of instruction at the Center for English as a Second Language at Southern Illinois University at Carbondale. Native languages represented in the subject pool included Spanish, Arabic, Chinese, Japanese, Korean, Malaysian, Thai, Greek, Tamil, Polish, French and Yoruba.

3 Instrumentation.

Data were elicited in the language laboratory with the following instruments:

The Wepman Auditory Discrimination Test consists of 40 word pairs equated for length; 30 of the pairs differ in a single phoneme, while 10 pairs are identical. The subject must indicate whether the spoken pair is the same or different.

Three tests from the Seashore Measure of Musical Talents were employed. In the Seashore test of pitch, the subject is presented 50

FL015450

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

J. Tomaras

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

pairs of tones, and the subject must determine whether the second tone is higher or lower in pitch than the first. The tones are at about 500 cycles and have a duration of 0.6 second each. Frequency differences in cycles between the tones in the pairs range from 2 to 17. In the Seashore test of loudness, the subject is presented 50 pairs of tones, and the subject must indicate for each pair whether the second tone is stronger or weaker than the first; the frequency is held constant at 440 cycles. The intensity differences in decibels between the tones in the pairs range from 0.5 to 4.0. In the Seashore test of rhythm, the subject is required to indicate whether a series of 30 pairs of rhythmic beats are the same or different.

The sentence repetition test involves memory; it is a measure of the immediate recall of sentence length verbal material. The test consists of 26 tape-recorded sentences of increasing length and complexity from "Mary agreed" to "the sun never reaches the ground through the overhead canopy of trees and vines." The subject must repeat each sentence immediately after two presentations. There is a twenty second interval between the second presentation and the next sentence. For scoring, one point was allowed for each sentence. To receive credit, the subject must repeat each word in the sentence; the words must be correctly sequenced, and all bound morphology must be present. For example, if a subject repeated "Mary agree" for the stimulus "Mary agreed," the subject receives no point for that sentence. The 26 sentences contained a total of 216 words. The longer sentences were included in the test due to Hammer's (1974) observation that listeners lose the exact wording of a sentence when only 16 syllables (for native speakers) intervened between the presentation of a sentence and its test. Pilot work for this research indicated that 7 to 8 syllables is the threshold for ESL subjects.

The Test of Spoken English (TSE) was used to evaluate the oral English speaking proficiency of the subject pool, and it consists of seven sections, each of which requires a different speaking activity. The following description is found in the Guide to SPEAK (1982:6-7):

> The first section is an unscored "warm-up" in which the examinee responds orally to a few brief biographical questions provided on the test tape.
>
> In the second section, the examinee is allowed time for preliminary silent reading of a passage of about 125 words and then is instructed to read the passage aloud. Scoring is based on pronunciation and overall clarity of speech.
>
> In the third section, the examinee is asked to complete a series of ten partial sentences in a way that conveys meaning and is grammatically correct.
>
> The fourth section of the test consists of six line drawings that tell a continuous story. After studying the drawings briefly, the examinee is asked to tell the story that is depicted, using past tense narration.
>
> In the fifth section, the examinee looks at a single line drawing and answers several spoken questions about the picture.
>
> The sixth section consists of a series of spoken questions intended to elicit relatively free and somewhat more lengthy

BEST COPY AVAILABLE

responses. Questions requiring both straightforward descriptions of common objects and fairly open-ended expressions of opinion are included. The linguistic quality and adequacy of communication, not the factual content of the responses, are at issue in scoring.

In the seventh and final section, the examinee sees a printed schedule, such as the outline for a course or a conference, and is asked to describe the schedule aloud, as though informing a group of listeners.

The TSE rating schedules are presented in Appendix A.

These instruments were chosen because previous research has shown that auditory factors and memory span are highly related to attained second and foreign language proficiency (e.g., Guilford and Christensen 1956; Carroll 1961; Holzinger and Harman 1938; Karlin 1942; Carroll 1962; Pimsleur, Stockwell and Comrey 1962; Lautenegger and Mueller 1964; Lautenegger, Mueller and Morahan 1965; Blickenstaff 1963; Suter 1976; Purcell and Suter 1981).

4 Analyses and Results

The means, standard deviations, and estimates of reliability for all measures are presented in Table 1. The correlation matrix is displayed in Table 2.

To determine whether one or more factors were manifested in the test data, we selected the principal factor solution with iterations. The criterion for the number of factors to be retained was an eigenvalue for each factor equal to or greater than 1. The principal factor solution with iterations is presented in Table 3.

We selected the varimax rotation option to make each factor define in a unique manner a distinct cluster of inter-correlated variables in the factor structure, provided that such clusters do indeed exist. If the unitary factor hypothesis is to be confirmed by this research, then only one such cluster should be found in the factor structure. If more than one distinct cluster of intercorrelated variables exist, this research will disconfirm the unitary competence hypothesis.

We wanted to be objective about how large, significant or salient a factor loading must be in the varimax rotated solution to merit attention. Following Gorsuch (1974), we established that with samples as small as ours, a factor loading of ± .64 is significant. Gorsuch suggests that a rough check on the minimum salient loading is provided by doubling the appropriate standard error. For our sample, the minimum significant correlation coefficient ($p<.01$) with an n of 50 is .32; therefore, loadings greater than .64 are considered salient or significant.

The varimax rotated factor matrix is presented as Table 4. Two factors were retained and rotated. All the speaking measures save for sentence repetition load significantly on Factor 1 which we will

BEST COPY AVAILABLE

refer to as a speaking or articulatory factor. We assume that the reason why the sentence repetition test did not significantly load on the articulatory factor is that it is also a measure of listening comprehension and short-term memory. We shall refer to Factor 2 as the auditory discrimination factor because there were moderate loadings by the different measures of auditory discrimination. Only one measure--the Seashore Loudness test--loaded significantly but by a hair-thin margin. The results of Table 4 seem to indicate that the unitary competence hypothesis was disconfirmed by this research. If there were support for the unitary competence hypothesis, we should have seen both the auditory and articulatory measures load significantly on one factor.

We must point out that these results should be interpreted with caution for the following reasons: (1) we believe that one should have at least 30 to 35 subjects per measure before submitting a data set to factor analytic solutions. In our case that would mean at least 270 subjects (we had 50); (2) there are doubtlessly more cautious and conservative methods of determining how high a factor loading must be to be significant, that is, we may have set the criterion factor loading too low.

For the reasons set out in the preceding paragraph we conducted two additional non-factor-analytic analyses. First, we corrected the correlation matrix for attenuation (Table 6). Then we partitioned the correlation matrix into three sections: (1) a cluster of intercorrelations of the listening discrimination measures---what we term a monotrait-heteromethod cluster. All the measures in this group purport to measure listening discrimination but employ different methods such as differences of loudness, pitch, phonemes, etc.; (2) a cluster of intercorrelations of the speaking or articulatory measures--a monotrait-heteromethod cluster; and (3) a cluster of intercorrelations of both the auditory and articulatory measures--a heterotrait-heteromethod cluster. The correlation coefficients for each cluster were averaged. As Table 5 indicates, the mean correlation for the articulatory measures (.723) is over twice as large as the mean for the auditory discrimination measures (.340) and almost seven times greater than the mean for the heterotrait-heteromethod cluster (.128). We believe that the data in Table 5 corroborate our factor-analytic results. If there were only one factor underlying the data, we would expect that the mean correlation for the heterotrait-heteromethod cluster would have been more in line with the other two means.

We did not submit these means to any statistical test of significance, because we are not sure if such a test would be appropriate. However, if such a test were deemed appropriate, we recommend that the correlation coefficients be transformed to a normal interval scale by using the Fisher Z transformation before they are averaged.

A further non-factor-analytic procedure involved determining the proportion of the reliable variance which is shared between pairs of variables. The procedure involves squaring the correlation "corrected" for unreliability:

BEST COPY AVAILABLE

$$r^2 = \left[\frac{r_{xy}}{\sqrt{r_{xx}\, r_{yy}}} \right]^2$$

where r_{xy} is the correlation coefficient obtained between the x and y variables

r_{xx} is the estimate of reliability of the x variable and

r_{yy} is the estimate of reliability of the y variable.

The proportions of shared reliable variance are reported in Table 6. We believe that the data displayed in Table 6 also corroborate the factor analysis results and interpretations. If there were a single unitary factor underlying these data, we would expect the shared reliable variance between auditory and articulatory measures to be as high or nearly as high as the variance between any two auditory variables or any two articulatory variables. Table 6 indicates that this is not the case.

A related issue--although a thorough treatment would go beyond the scope of this paper--is the amount of variance which is attributable to the trait and the amount of variance which is attributable to the test method. Bachman and Palmer's (1981) research indicates that method effects can account for substantial amounts of variance. The question of whether the trait/method variance proportion is 50/50, 67/33, 25/75, etc., remains open.

In summary, the results of this investigation suggest that there are two distinct factors present in the data analyzed--an auditory factor and an articulatory factor. With more data elicitation and more sophisticated analysis such as multitrait-multimethod validation and confirmatory factor analysis, one might find evidence for a method factor. The present analyses and interpretations tend to disconfirm the unitary competence hypothesis proposed by Oller (1979) and Oller and Hinofotis (1980) and seem to be more in accordance with the multiple-factor models proposed by Upshur and Homburg (1983) and Bachman and Palmer (1981).

BEST COPY AVAILABLE

APPENDIX A

TSE Rating Schedules

Pronunciation

0 - Frequent phonemic errors and foreign stress and intonation patterns that cause the speaker to be unintelligible

1 - Frequent phonemic errors and foreign stress and intonation patterns that cause the speaker to be occasionally unintelligible

2 - Some consistent phonemic errors and foreign stress and intonation patters but speaker is intelligible

3 - Occasional nonnative pronunciation errors but speaker is always intelligible

Grammar

0 - Virtually no grammatical or syntactical control except in simple stock phrases

1 - Some control of basic grammatical constructions but with major and/ or repeated errors that interfere with intelligibility

2 - Generally good control in all constructions with grammatical errors that do not interfere with overall intelligibility

3 - Sporadic minor grammatical errors that could be made inadvertantly by native speakers

Fluency

0 - Speech is so halting and fragmentary or has such a nonnative flow that intelligibility is virtually impossible

1 - Numerous nonnative pauses and/or a nonnative flow that interferes with intelligibility

2 - Some nonnative pauses but with a more nearly native flow so that the pauses do not interfere with intelligibility

3 - Speech is as smooth and as effortless as that of a native speaker

Comprehensibility

0 - Overall comprehensibility too low in even the simplest type of speech

BEST COPY AVAILABLE

1 - Generally not comprehensible due to frequent pauses and/or rephrasing, pronunciation errors, limited grasp of vocabulary, or lack of grammatical control

2 - Comprehensible with errors in pronunciation, grammar, choice or vocabulary items, or infrequent pauses or rephrasing

3 - Completely comprehensible in normal speech with occasional grammatical or pronunciation errors

BEST COPY AVAILABLE

Table 1

Means, standard deviations, and estimates of reliability

Test	Number of subjects	Number of items/points	Mean	SD	Type of reliability	Estimate of reliability
Wepman Auditory Discrimination	50	40	36.54	1.876	KR-20	.320
Seashore Pitch	50	50	31.02	7.844	KR-20	.867
Seashore Loudness	50	50	41.26	9.167	KR-20	.938
Seashore Rhythm	50	30	24.46	4.234	KR-20	.803
Sentence Repetition	50	26	9.64	3.768	Interrater	.910
TSE Pronunciation	50	3	1.63	.468	Interrater	.935
TSE Grammar	50	3	1.65	.522	Interrater	.847
TSE Fluency	50	3	1.70	.492	Interrater	.727
TSE Comprehensibility	50	300	171.20	50.485	Interrater	.826

BEST COPY AVAILABLE

Table 2

Correlation matrix

	1	2	3	4	5	6	7	8	9
1. Wepman Auditory Discrimination	1.00	.15 p=.144	.24 p=.045	.13 p=.185	.39 p=.003	-.04 p=.380	.05 p=.367	-.08 p=.289	-.05 p=.374
2. Seashore Pitch		1.00	.43 p=.001	.21 p=.073	-.001 p=.497	-.09 p=.264	-.16 p=.126	-.05 p=.372	-.14 p=.166
3. Seashore Loudness			1.00	.29 p=.019	.25 p=.038	.15 p=.150	.11 p=.221	.23 p=.056	.15 p=.150
4. Seashore Rhythm				1.00	.40 p=.002	.33 p=.010	.14 p=.157	.18 p=.109	.27 p=.029
5. Sentence Repetition					1.00	.44 p.001	.38 p=.003	.42 p=.001	.36 p=.005
6. TSE Pronunciation						1.00	.75 p=.000	.77 p=.000	.89 p=.000
7. TSE Grammar							1.00	.62 p=.000	.89 p=.000
8. TSE Fluency								1.00	.74 p=.000
9. TSE Comprehensibility									1.00

BEST COPY AVAILABLE

Table 3

Principal factor solution with iterations

	Factor 1	Factor 2	Communality
Wepman Auditory Discrimination	.05633	.41957	.17922
Seashore Pitch	-.06361	.52877	.28365
Seashore Loudness	.24894	.61374	.43864
Seashore Rhythm	.33772	.40003	.27408
Sentence Repetition	.51637	.34792	.38768
TSE Pronunciation	.92409	-.09647	.86325
TSE Grammar	.82964	-.17495	.71892
TSE Fluency	.77699	-.04695	.60592
TSE Comprehensibility	.96015	.19451	.95972

Factor	Eigenvalue	Percent of Variability	Cumulative Percent
1	3.51722	74.7	74.7
2	1.19322	25.3	100.0

BEST COPY AVAILABLE

Table 4

Varimax rotated factor matrix

	Factor 1	Factor 2
Wepman Auditory Discrimination	-.02372	.42267
Seashore Pitch	-.16208	.50732
Seashore Loudness	.12887	.64964*
Seashore Rhythm	.25632	.45649
Sentance Repetition	.44158	.43896
TSE Pronunciation	.92572*	.07933
TSE Grammar	.84775*	-.01554
TSE Fluency	.77193*	.10026
TSE Comprehensibility	.97960*	-.01016

*significant (salient) factor loading

BEST COPY AVAILABLE

Table 5

Correlation matrix corrected for attenuation

	1	2	3	4	5	6	7	8	9	
1. Wepman Auditory Discrimination		.285	.438	.256	.723	-.073	.96	-.166	-.097	heterotrait heteromethod $\bar{r}$=.128
2. Seashore Pitch	monotrait-heteromethod $\bar{r}$=.340		.477	.252	.001	-.100	-.187	-.063	-.165	
3. Seashore Loudness				.344	.271	.160	.123	.278	.170	
4. Seashore Rhythm					.468	.381	.169	.236	.332	
5. Sentence Repetition						.477	.433	.517	.415	monotrait-heteromethod $\bar{r}$=.723
6. TSE Pronunciation							.844	.934	.910	
7. TSE Grammar								.789	.956	
8. TSE Fluency									.955	
9. TSE Comprehensibility										

Table 6

Shared reliable variance between variables

	1	2	3	4	5	6	7	8	9	
1. Wepman Auditory Discrimination		.081	.661	.066	.523	.005	.009	.027	.009	heterotrait-heteromethod $\bar{X}$=.071
2. Seashore Pitch			.227	.063	.000	.010	.035	.004	.027	
3. Seashore Loudness	monotrait-heteromethod $\bar{X}$=.279			.578	.073	.026	.015	.077	.029	
4. Seashore Rhythm					.219	.146	.029	.066	.110	
5. Sentence Repetition						.227	.187	.267	.172	
6. TSE Pronunciation							.712	.872	.828	monotrait-heteromethod $\bar{X}$=.572
7. TSE Grammar								.822	.916	
8. TSE Fluency									.912	
9. TSE Comprehensibility										

BEST COPY AVAILABLE

REFERENCES

Bachman, L. and Palmer, A. 1981. "The construct validation of the FSI oral interview". Language Learning 31: 67-86.

Blickenstaff, C. 1963. "Musical talents and foreign language learning ability." Modern Language Journal XLVII: 359-363.

Carroll, J. 1941. "A factor analysis of verbal abilities". Psychometrika 6: 279-307.

Carroll, J. 1962. "The prediction of success in intensive foreign language training". In Glaser, R. (ed). Training research and education. Pittsburgh, Pennsylvania: University of Pittsburgh Press.

Educational Testing Service. 1982. Guide to SPEAK. Princeton, New Jersey: Author.

Educational Testing Service. 1982. Test of spoken English. Princeton, New Jersey: Author.

Gorsuch, R. 1974. Factor analysis. Philadelphia, Pennsylvania: W. B. Saunders Company.

Guilford, J. and Christensen, P. 1956. "A factor-analytic study of verbal fluency". University of Southern California Psychology Laboratory Report No. 17. Los Angeles: University of Southern California.

Holzinger, K. and Harmon, J. 1938. "Comparison of two factor analyses". Psychometrika 3: 45-60.

Karlin, J. 1942. "The factorial isolation of the primary auditory abilities". Unpublished Ph.D. dissertation, University of Chicago.

Leutenegger, R. and Mueller, T. 1964. "Auditory factors and the acquisition of French language mastery". Modern Language Journal XLVII: 141-146.

Leutenegger, R., Mueller, T., and Mershon, I. 1965. "Auditory factors in foreign language acquisition". Modern Language Journal XLIX: 22-31.

Oller, J. Jr. 1979. Language tests at school: a pragmatic approach. London: Longman.

Oller, J. Jr. and Hinofotis, F. 1980. "Two mutually exclusive hypotheses about second language ability: indivisible or partially divisible competence". In Oller, J. and Perkins, K., (eds). Research in language testing. Rowley, Massachusetts: Newbury House.

BEST COPY AVAILABLE

Pimsleur, P., Stockwell, R., and Comrey, A. 1962. "Foreign language learning ability". Journal of Educational Psychology 53: 15-26.

Purcell, E. and Suter, R. 1981. "Predictors of pronunciation accuracy: a reexamination". Language Learning 30: 271-288.

Seashore, C., Lewis, D., and Saetveit, J. 1939. Seashore measures of musical talents. New York: The Psychological Corporation.

Suter, R. 1976. "Predictor of pronunciation accuracy in second language learning". Language Learning 26: 233-253.

Upshur, J. and Homburg, T. 1983. "Some relations among language tests at successive ability levels". In Oller, J. Jr. (ed). Issues in language testing research. Rowley, Massachusetts: Newbury House.

Wanner, E. 1974. On remembering, forgetting, and understanding sentences. The Hague: Mouton.

Wepman, J. 1973. Auditory discrimination test. Los Angeles: Western Psychological Services.

BEST COPY AVAILABLE